Hacettepe Journal of
Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 53 (3) (2024), 851 – 878
DOI : 10.15672/hujms.1312795

Research Article

Study of a two types of general heterogeneous service queueing system in a single server with optional repeated service and feedback queue

Snigdha Mahanta*[1], Nitin Kumar[2], Gautam Choudhury[1]

[1] *Department of Mathematical Sciences, Institute of Advanced Study in Science and Technology, Guwahati-781035, India*

[2] *Department of Mathematics, Indian Institute of Technology Jammu, Jammu-181221, India*

Abstract

This paper addresses a model on a single server queue and two service representatives. After a customer is served, he/she has the three options: opting for receive the same service again (re-service), joining as a new customer for another regular service (feedback), or leaving the service system altogether. To ensure the queueing system is Markovian, we introduce an additional variable (supplementary variable) and using this approach, we derive the explicit distribution of queue size at random and departure epochs. Additionally, we determine the distribution of response time, inter-departure time, and busy period. By using the embedded Markov chain technique we have also derived the queue size distribution at departure epoch. We have also presented the cost analysis of the model with some numerical examples. The numerical illustration validates our findings and provides valuable insights into the queuing system.

Mathematics Subject Classification (2020). 60K25, 90B22

Keywords. Bernoulli feedback, busy period, queueing system, re-service

1. Introduction

In a queueing system where a single server offers two different types of services, it is commonly referred to as a single-server queueing system with two types of heterogeneous service. The initial theory for this type of queueing model was developed by [2]. They made the assumptions that the service times of customers are independent and follow an exponential distribution. Additionally, the server takes a single vacation, which is influenced by Bernoulli schedules, where the vacation period follows an exponential distribution. They obtained explicit steady-state results for the probability generating functions of the queue size and system size, along with other performance measures of the system. Madan et al. [16] extended this research by studying a single server queue with batch arrivals and two types of heterogeneous service, where the general service time distributions differ. They

*Corresponding Author.
Email addresses: snigdhamahanta11@gmail.com (S. Mahanta), nitinkumar7276@gmail.com (N. Kumar), gc.iasst@gmail.com (G. Choudhury)
Received: 11.06.2023; Accepted: 26.03.2024

also introduced the possibility of customers opting for re-service after completing their initial service. They obtained steady-state probability generating function for the queue size and system size, as well as important performance measures such as the average number of customers and the average waiting time in the queue and system. Another similar model was investigated by [17], where they assumed general vacation time and Bernoulli schedule for server vacation under a single vacation policy. They derived the steady-state queue size distribution, the mean busy period of the server, and other performance measures of the system. Kalita and Choudhury [6] investigated an $M/G/1$ queue with two types of general heterogeneous service and optional repeated service subject to servers breakdown and delayed repair. Also, in this area, (that is, in the area of the two types of general heterogeneous service) can be found in the study conducted by [12] where they investigated the two types of general heterogeneous service on $M/G/1$ queue with delayed repair under randomized vacation policy. Moreover, Begum and Choudhury [3] studied $M/G/1$ queue with two types of general heterogeneous service with Bernoulli vacation and server breakdown, the server after completion of a service is allowed to take in a single vacation under the Bernoulli schedule.

In the context of Bernoulli feedback, when a customer is unsatisfied with their service, they have the option to join the queue again for service with a certain probability. Takacs [24] provides detailed insights into the Bernoulli feedback mechanism in his book. Takacs [25] investigated a queueing model of $M/G/1$ type with Bernoulli feedback and determined the distribution of queue size as well as the first two moments of the distribution function representing the total time spent by a customer in the system. Rege [23] studied the $M/G/1$ queue with Bernoulli feedback, offering a proof based on branching process to simplify the derivation of certain performance measures. Rege [23] also explained the significance of these performance measures in practical terms. Choudhury and Paul [7] extended the Bernoulli feedback mechanism to a two-phase of heterogeneous service $M/G/1$ queue. They derived the distribution of queue size and service completion epochs, as well as the distribution of response time and busy period. Kumar et al. [14] examined a non-Markovian feedback single-server retrial queue with collisions and general retrial times. They used the supplementary variable technique to obtain the steady-state joint distribution of the server's state and the orbit length. Important performance measures and stability conditions of the system were also provided. Choi et al. [5] considered an $M/G/1$ queue with Bernoulli feedback and multiple classes of customers, each with different arrival rates, service time distributions, and feedback parameters. They obtained the joint probability generating function of the system size for each class, allowing them to calculate the moments of the system size and total response time for each customer class. Mahanta and Choudhury [18] studied feedback queue on $M/\binom{G_1}{G_2}/1$ with vacation. Using the supplementary variable technique they obtained the probability generating function of the queue size distribution at random epoch and departure epoch. They also obtained mean queue size at random epoch, mean waiting time and mean busy period of the system. Jain and Kaur [11] studied Bernoulli feedback on $M^x/G/1$ unreliable server retrial bulk queue with multiphase optional service incorporating the features of balking, Bernoulli vacation. For evaluating the queue size distribution and other system performance metrics, supplementary variable technique is used. The approximate solutions for the steady state probabilities and waiting time are suggested using maximum entropy principle (MEP). For the discrete-time queueing models, in this direction, one can refer to [15] and [28]. GnanaSekar and Kandaiyan [10] studied the dynamics of feedback within a single server retrial queueing system with delayed repair under a working vacation policy, where customers are allowed to balk and renege in some situations. Utilizing the supplementary variable technique, they derived the steady state probability generating function for both system size and orbit size. Furthermore, they conducted an analysis of system parameter impacts through numerical examples. Khan and Paramasivam [13] investigated an

$M/M/1/N$ encouraged arrival queue with feedback, balking, and the management of reneged clients. They examined a quality control policy for the Markovian model using an iterative approach up to the n^{th} customer in the system. Additionally, they determined performance measures such as the expected number of units in the system and queue, average number of occupied services, and expected waiting time in both the system and queue. Melikov et al. [21] delved into a Markov model of a queuing-inventory system encompassing primary, retrial, and feedback customers. They utilized the matrix-geometric method to compute steady-state probabilities. Mahanta et al. [19] investigated Markov modulated Poisson input with feedback queue under N - policy. GnanaSekar and Kandaiyan [9] studied feedback retrial queue with two dependent phases of service under Bernoulli working vacation. The service times for the two stages are often independent in normal queueing frameworks. The first phase service time has an impact on the second phase service time. In order to determine the steady-state probabilities and probability-generating function for the different states, the supplementary variable technique was utilized. Furthermore, Niranjan et al. [22] considered an essential two-phase bulk service, immediate Bernoulli feedback for customers, and renewal service time of the first essential service for the bulk arrival and bulk service queueing model. They investigated the probability-generating function of the queue size at any time. Additionally, they conducted an optimum cost analysis to minimize total average cost, with practical applications in existing data transmission and processing in LTE-A networks using the DRX mechanism.

The novelty of this paper lies in the fact that, to the best of the authors' knowledge, no previous study in the literature has examined a single server $M/G/1$ queueing system with two types of general heterogeneous service, repeated service, and Bernoulli feedback in a single aspect. Motivated by this gap, the authors aim to analyze such a model. The paper analyzes two types of service representatives and assumes a first-come-first-served service discipline, allowing customers to choose their desired service system. Customers can either opt for the same type of service again or join as new customers to receive another regular service. The model is analyzed using the supplementary variable technique, where the elapsed service time serves as the supplementary variable. The steady-state queue size distribution at arbitrary epoch is obtained, along with the distribution of various time durations such as inter-departure time, response time, and busy period for arriving customers. Numerical results are presented to demonstrate the applicability of the model.

The structure of the paper is as follows: Section 2 describes the mathematical model. Complete theoretical analysis of the model in the steady state is presented in Section 3. In Section 4 the embedded Markov chain result for the model is presented and in Section 5 the cost analysis is carried out. Section 6 illustrates a simulation study to validate the analytical results, followed by practical applications of the model is briefly discussed in Section 7. Finally, the paper concludes in Section 8.

2. Model description

In this Section, we describe the mathematical model which is stated below:

The customers arrive to the queueing system according to the Poisson process, that is, the inter-arrival times of the customers are independent and exponentially distributed where λ represents the arrival rate of the customers.

The system consists of a single server that offers two different types of services, namely the first type of service (FTS) denoted as B_1, and the second type of service (STS) denoted as B_2. Further, $b_j(x)$, $B_j(x)$, and $B_j^*(s)$ be the probability density function (p.d.f), probability distribution function (PDF), Laplace transform of B_j-th service, respectively, and the corresponding moments are denoted by $\beta_j^{(k)}$, $k \geq 1$.

The customer also has the opportunity to select any one of the representative services - either he can select the FTS with probability p_1 or he can select the STS with probability

$p_2(=1-p_1)$. The total service time needed by a unit for completing the service cycle in the following manner.

$$B = \begin{cases} B_1, & \text{with probability } p_1 \\ B_2, & \text{with probability } p_2. \end{cases}$$

Once a service system is selected, the Laplace-Steiltjes transform (LST) for the total duration of the service time can be determined as

$$B^*(s) = p_1 B_1^*(s) + p_2 B_2^*(s).$$

Once a customer finishes receiving any of the representative services, they are given the choice to either repeat the same service one more time (but only once), or to exit from the queueing system. Let the probability of repeating the same type of service be f_j whereas the probability of not repeating the service and goes away from the system be $1-f_j, j=1,2$. Further, let $F_j, I_j(x)$ and $I_j^*(s)$ denote the j^{th} type of repeated service time, corresponding distribution function and Laplace transform respectively. Also $\gamma_j^{(k)}, k \geq 1$ denotes the finite moments of the repeated service time. Therefore, the modified service is

$$B_j \quad = \quad \begin{cases} B_j + F_j, & \text{with probability } f_j \\ B_j, & \text{with probability } 1 - f_j. \end{cases}$$

The LST $B_j^*(s)$ of repeated service F_j for $j = 1, 2$ is

$$B_j^*(s) = f_j B_j^*(s) I_j^*(s) + (1 - f_j) B_j^*(s).$$

Therefore the LST of the modified service time is given by

$$B^*(s) = (f_1 B_1^*(s) + (1 - f_1)) p_1 B_1^*(s) + (f_2 B_2^*(s) + (1 - f_2)) p_2 B_2^*(s). \tag{2.1}$$

And the first two raw moments of the modified service time are given by

$$\beta_1 \quad = \quad (1 + f_1)\, p_1 \beta_1^{(1)} + (1 + f_2)\, p_2 \beta_2^{(1)}, \tag{2.2}$$

$$\beta_2 \quad = \quad \left(\beta_1^{(2)} + f_1 \left(\beta_1^{(2)} + 2\left(\beta_1^{(1)}\right)^2\right)\right) p_1 + \left(\beta_2^{(2)} + f_2 \left(\beta_2^{(2)} + 2\left(\beta_2^{(1)}\right)^2\right)\right) p_2. \tag{2.3}$$

If for a certain reason a customer is not satisfied with its service then the customer can also immediately join or move on of the tail of the original queue to receive another regular service with a certain probability Θ, or they can move away from the system with probability $\Theta^{'}$ $(=1-\Theta)$. Let Γ and $\Omega(x)$ denote the Bernoulli feedback and corresponding general distribution function, respectively.

3. Analysis of the model

Let us define the following random variable at time t:

- $N_q(t) :=$ Queue size.
- $B_j^0(t) :=$ Elapsed j^{th} representative service.
- $F_j^0(t) :=$ Elapsed j^{th} type of re-service.
- $Y(t) :=$ State of the system.

The state of the system $(Y(t))$ at time t is given as

$$Y(t) \quad = \quad \begin{cases} 0, & \text{if the system is idle at time } t \\ 1, & \text{if the system is occupied by the FTS at time } t \\ 2, & \text{if the system is occupied by the STS at time } t \\ 3, & \text{if the system is occupied by the FRS at time } t \\ 4, & \text{if the system is occupied by the SRS at time } t, \end{cases} \tag{3.1}$$

where we assign FRS and SRS to indicate the first and second type of service respectively, that are subject to repetition. The supplementary variables $B_1^0(t)$, $B_2^0(t)$, $F_1^0(t)$, $F_2^0(t)$ are introduced in order to obtain a bivariate Markov process $\{N_Q(t), \chi(t)\}$, where

$$\chi(t) = \begin{cases} 0, & \text{if } Y(t) = 0 \\ B_1^0(t), & \text{if } Y(t) = 1 \\ B_2^0(t), & \text{if } Y(t) = 2 \\ F_1^0(t), & \text{if } Y(t) = 3 \\ F_2^0(t), & \text{if } Y(t) = 4. \end{cases} \quad (3.2)$$

Let us define the following probabilities:

$$\begin{aligned} V_0(t) &= Prob[N_q(t) = 0, \chi(t) = 0], \\ A_{n,j}(x;t)dx &= Prob[N_q(t) = n,\ \chi(t) = B_j^0(t);\ x < B_j^0(t) \leq x + dx],\ x > 0, n \geq 0,\ j = 1, 2, \\ B_{n,j}(x;t)dx &= Prob[N_q(t) = n, \chi(t) = F_j^0(t);\ x < F_j^0(t) \leq x + dx],\ x > 0, n \geq 0,\ j = 1, 2. \end{aligned}$$

We have $B_j^0(0) = 0,\ B_j^0(\infty) = 1,\ F_j^0(0) = 0,\ F_j^0(\infty) = 1;\ j = 1, 2$ and $B_j(x)$ is continuous at $x = 0$ such that,

$$\xi_j(x) = \frac{b_j(x)}{1 - B_j(x)}$$

are the first order differential (hazard rate) function of B_j and therefore,

$$b_j(\Omega) = \xi_j(\Omega) e^{-\int_0^\Omega \xi_j(x)dx}, \quad j = 1, 2.$$

Further, the steady-state probability generating functions (PGF's) are given as

$$A_j(x, z) = \sum_{n=0}^{\infty} A_{n,j}(x) z^n, \qquad A_j(z) = \sum_{n=0}^{\infty} A_{n,j} z^n,\ j = 1, 2, \quad (3.3)$$

$$B_j(x, z) = \sum_{n=0}^{\infty} B_{n,j}(x) z^n, \qquad B_j(z) = \sum_{n=0}^{\infty} B_{n,j} z^n,\ j = 1, 2, \quad (3.4)$$

where

$$\begin{aligned} A_{n,j}(x) &= \lim_{t\to\infty} A_{n,j}(x;t), \qquad A_{n,j} = \int_0^\infty A_{n,j}(x)dx,\ j = 1, 2, \\ B_{n,j}(x) &= \lim_{t\to\infty} B_{n,j}(x;t), \qquad B_{n,j} = \int_0^\infty B_{n,j}(x)dx,\ j = 1, 2, \end{aligned}$$

and

$$V_0 = \lim_{t\to\infty} V_0(t).$$

The state transition diagram of this queuing system is depicted in Figure 1, illustrating the system's different states using a 2-tuple, namely (α, η). The value of α represents the number of customers in the system at a given time t, while η indicates the state of the system at that time. The variable α can assume values ranging from 0 onwards, indicating the presence of 0, 1, 2, and so on customers, while η can assume values of 0, 1, 2, 3, 4, or 5. In this system, state 0 represents the idle period, state 1 represents the FTS period, state 2 represents the FRS period, state 3 represents the STS period, state 4 represents the SRS period, and state 5 signifies the Benoulli feedback time.

Now using the argument of [8], we obtain the following steady-state equations of the system.

$$\begin{aligned} \lambda V_0 &= (1 - f_1)\Theta' \int_0^\infty A_{0,1}(x)\xi_1(x)dx + (1 - f_2)\Theta' \int_0^\infty A_{0,2}(x)\xi_2(x)dx \\ &+ \Theta' \int_0^\infty B_{0,1}(x)\xi_1(x)dx + \Theta' \int_0^\infty B_{0,2}(x)\xi_2(x)dx. \quad (3.5) \end{aligned}$$

$$\frac{d}{dx}A_{n,j}(x) + (\lambda + \xi_j(x))A_{n,j}(x) = \lambda A_{n-1,j}(x), \quad n \geq 0;\ j = 1, 2, \tag{3.6}$$

$$\frac{d}{dx}B_{n,j}(x) + (\lambda + \xi_j(x))B_{n,j}(x) = \lambda B_{n-1,j}(x), \quad n \geq 0;\ j = 1, 2. \tag{3.7}$$

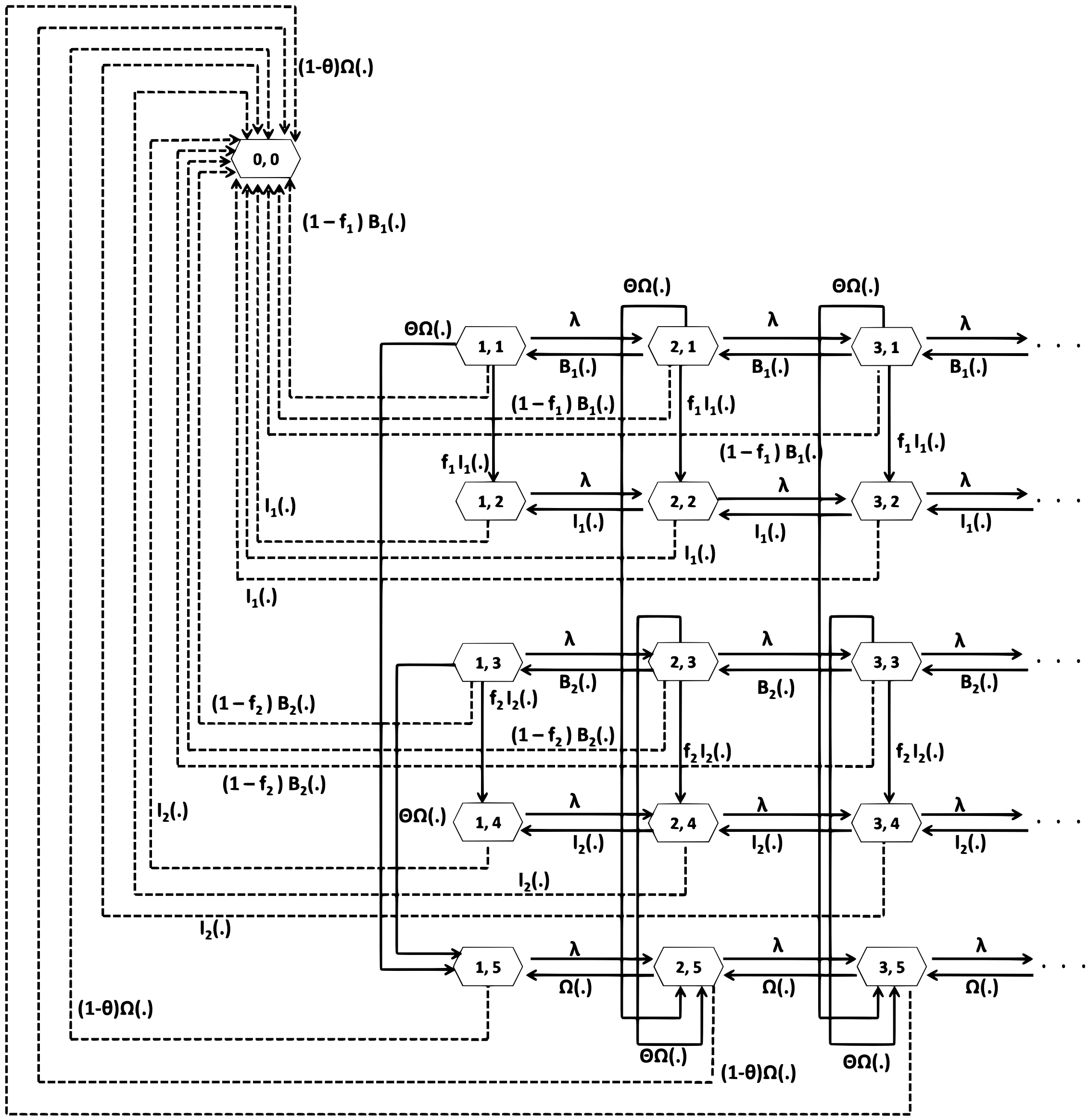

Figure 1. State transition diagram of the model.

The boundary conditions for solving the aforementioned steady-state equations are

$$\begin{aligned} A_{n,1}(0) &= \sum_{j=1}^{2}(1-f_j)p_1\left(\Theta\int_0^\infty A_{n,j}(x)\xi_j(x)dx + \Theta'\int_0^\infty A_{n+1,j}(x)\xi_j(x)dx\right) \\ &+ \sum_{j=1}^{2}p_1\left(\Theta\int_0^\infty B_{n,j}(x)\xi_j(x)dx + \Theta'\int_0^\infty B_{n+1,j}(x)\xi_j(x)dx\right),\ n \geq 0,\ j = 1, 2, \end{aligned} \tag{3.8}$$

$$\begin{aligned} A_{0,1}(0) &= \sum_{j=1}^{2}(1-f_j)p_1\left(\Theta\int_0^\infty A_{0,j}(x)\xi_j(x)dx+\Theta'\int_0^\infty A_{1,j}(x)\xi_j(x)dx\right) \\ &+ \sum_{j=1}^{2}p_1\left(\Theta\int_0^\infty B_{0,j}(x)\xi_j(x)dx+\Theta'\int_0^\infty B_{1,j}(x)\xi_j(x)dx\right)+\lambda V_0 p_1,\ j=1,2, \end{aligned} \tag{3.9}$$

$$\begin{aligned} A_{n,2}(0) &= \sum_{j=1}^{2}(1-f_j)p_2\left(\Theta\int_0^\infty A_{n,j}(x)\xi_j(x)dx+\Theta'\int_0^\infty A_{n+1,j}(x)\xi_j(x)dx\right) \\ &+ \sum_{j=1}^{2}p_2\left(\Theta\int_0^\infty B_{n,j}(x)\xi_j(x)dx+\Theta'\int_0^\infty B_{n+1,j}(x)\xi_j(x)dx\right),\ n\geq 0,\ j=1,2, \end{aligned} \tag{3.10}$$

$$\begin{aligned} A_{0,2}(0) &= \sum_{j=1}^{2}(1-f_j)p_2\left(\Theta\int_0^\infty A_{0,j}(x)\xi_j(x)dx+\Theta'\int_0^\infty A_{1,j}(x)\xi_j(x)dx\right) \\ &+ \sum_{j=1}^{2}p_2\left(\Theta\int_0^\infty B_{0,j}(x)\xi_j(x)dx+\Theta'\int_0^\infty B_{1,j}(x)\xi_j(x)dx\right)+\lambda V_0 p_2,\ j=1,2, \end{aligned} \tag{3.11}$$

$$B_{n,j}(0) = f_j\int_0^\infty A_{n,j}(x)\xi_j(x)dx,\ j=1,2;\ n\geq 0. \tag{3.12}$$

Now using Equations (3.5)-(3.12), we obtain the steady state PGF of the queue size distribution which is given in the next section.

3.1. The PGF of the queue size distribution at random epoch

In this section, we provide the Probability Generating Function (PGF) for the distribution of the queue size, along with the condition required for the system to remain stable.

Denoting the PGF of the queue size, regardless of the service type, as $A_q(z)$, we solve Equations (3.5)-(3.12), and then substitute the results into Equations (3.3)-(3.4) and finally we derive the following results.

$$A_j(z) = \frac{\left(B_j^*(\lambda-\lambda z)-1\right)p_jV_0}{z-(z\Theta+\Theta')\Big(((1-f_1)+f_1B_1^*(\lambda-\lambda z))p_1B_1^*(\lambda-\lambda z) + ((1-f_2)+f_2B_2^*(\lambda-\lambda z))p_2B_2^*(\lambda-\lambda z)\Big)},\ j=1,2, \tag{3.13}$$

$$B_j(z) = \frac{\left(B_j^*(\lambda-\lambda z)-1\right)f_jp_jB_j^*(\lambda-\lambda z)V_0}{z-(z\Theta+\Theta')\Big(((1-f_1)+f_1B_1^*(\lambda-\lambda z))p_1B_1^*(\lambda-\lambda z) + ((1-f_2)+f_2B_2^*(\lambda-\lambda z))p_2B_2^*(\lambda-\lambda z)\Big)},\ j=1,2, \tag{3.14}$$

where $B_j^*(\lambda-\lambda z)=\int\limits_0^\infty e^{-(\lambda-\lambda z)x}dB_j(x)$.

Now using $A_q(z)=A_1(z)+A_2(z)+B_1(z)+B_2(z)$, we get

$$A_q(z) = \frac{\Big((B_1^*(\lambda-\lambda z)-1)(1+f_1B_1^*(\lambda-\lambda z))p_1 + (B_2^*(\lambda-\lambda z)-1)(1+f_2B_2^*(\lambda-\lambda z))p_2\Big)V_0}{z-(z\Theta+\Theta')\Big(((1-f_1)+f_1B_1^*(\lambda-\lambda z))p_1B_1^*(\lambda-\lambda z) + ((1-f_2)+f_2B_2^*(\lambda-\lambda z))p_2B_2^*(\lambda-\lambda z)\Big)}. \tag{3.15}$$

Now taking the limit $z \to 1$ in Equation (3.15), we get the steady state probability that the server is busy irrespective of the type of service as

$$A_q(1) = \lim_{z\to 1} A_q(z) = \frac{p_1 V_0 \lambda \beta_1^{(1)}(1+f_1) + p_2 V_0 \lambda \beta_2^{(1)}(1+f_2)}{\Theta' - p_1\lambda\beta_1^{(1)}(1+f_1) - p_2\lambda\beta_2^{(1)}(1+f_2)}. \tag{3.16}$$

Further, from the normalizing condition $V_0 + A_q(1) = 1$, we get the steady-state probability that the system is empty as

$$V_0 = 1 - \frac{p_1\lambda\beta_1^{(1)}(1+f_1) + p_2\lambda\beta_2^{(1)}(1+f_2)}{\Theta'}, \tag{3.17}$$

and the steady-state probability that the system is in idle state as

$$\rho = p_1\lambda(1+f_1)\beta_1^{(1)} + p_2\lambda(1+f_2)\beta_2^{(1)}. \tag{3.18}$$

Now from Equation (3.17), we have the condition that $\frac{\rho}{\Theta'} < 1$, which is the condition of stability necessary for the steady state solution to exist, see pp. 303 in [20].

Theorem 3.1. *The average queue size (M_q) under steady state is given by*

$$M_q = \frac{\begin{array}{c} 2f_1p_1\lambda^2\beta_1^{(2)} + (1-f_1)p_1\left[\lambda^2\beta_1^{(2)} + \lambda\beta_1^{(1)}\right] \\ + 2f_2p_2\lambda^2\beta_2^{(2)} + (1-f_2)p_2\left[\lambda^2\beta_2^{(2)} + \lambda\beta_2^{(1)}\right] \end{array}}{2\left[1 - \Theta - p_1\lambda\beta_1^{(1)}(1+f_1) - p_2\lambda\beta_2^{(1)}(1+f_2)\right]}. \tag{3.19}$$

Proof. Differentiating Equation (3.15) with respect to z and then setting $z = 1$, i.e., $M_q = \lim_{z\to 1} \frac{d}{dz} A_q(z)$, we obtain the required result. □

Remark 3.2. If W_q denotes the steady state average waiting time in the queue, then this can be obtained by the following relation.

$$W_q = \frac{M_q}{\lambda} \tag{3.20}$$

Theorem 3.3. *Under the stability condition $\frac{\rho}{\Theta'} < 1$, we can obtain the probabilities of different system states as follows:*

- **(A):** *The probability of the server being busy in the FTS is given by,* $P_{B_1} = \frac{\lambda p_1 \beta_1^{(1)}}{\Theta'}$.
- **(B):** *The probability of the server being busy in STS is given by,* $P_{B_2} = \frac{\lambda p_2 \beta_2^{(1)}}{\Theta'}$.
- **(C):** *The probability of the server being busy in FRS is given by,* $P_{F_1} = \frac{\lambda p_1 f_1 \beta_1^{(1)}}{\Theta'}$.
- **(D):** *The probability of the server being busy in SRS is given by,* $P_{F_2} = \frac{\lambda p_2 f_2 \beta_2^{(1)}}{\Theta'}$.

Proof. Taking limit $z \to 1$ in $A_j(z)$ and $B_j(z)$, we obtain the required results. □

3.2. PGF of the queue size distribution at departure epoch

Let h_i be the probability that there are i (≥ 0) customers in the queue at a departure epoch. Further, let $H(z)$ be the corresponding PGF, i.e., $H(z) = \sum_{i=0}^{\infty} h_i z^i$.

Theorem 3.4. *Under the condition* $\frac{\rho}{\Theta'} < 1$, *the PGF of queue size distribution at departure epoch* $H(z)$ *is given by*

$$H(z) = \frac{\begin{array}{l}\Big(\Theta' - p_1\lambda\beta_1^{(1)}(1+f_1) - p_2\lambda\beta_2^{(1)}(1+f_2)\Big)(\Theta z+\Theta')(1-z)\Big(((1-f_1) \\ \quad + f_1B_1^*(\lambda-\lambda z))p_1B_1^*(\lambda-\lambda z) + ((1-f_2)+f_2B_2^*(\lambda-\lambda z))p_2B_2^*(\lambda-\lambda z)\Big)\end{array}}{\begin{array}{l}z-(\Theta z+\Theta')\Big(((1-f_1)+f_1B_1^*(\lambda-\lambda z))p_1B_1^*(\lambda-\lambda z) \\ \qquad + ((1-f_2)+f_2B_2^*(\lambda-\lambda z))p_2B_2^*(\lambda-\lambda z)\Big)\end{array}}. \tag{3.21}$$

Proof. Based on the empirical evidence presented in [29], we can firmly state that if a customer observes i customers in the queue right after finishing their service, it means that there were exactly i customers in the queue right before their service ended. To represent the probability of having i customers in the queue at the service completion epoch, and by denoting it as h_i where $i = 0, 1, 2, \dots,$ we can express as follows:

$$\begin{aligned} h_0 &= T(1-f_1)\Theta'\int_0^\infty \xi_1(x)A_{0,1}(x)dx + T(1-f_2)\Theta'\int_0^\infty \xi_2(x)A_{0,2}(x)dx \\ &+ T\Theta'\int_0^\infty \xi_1(x)B_{0,1}(x)dx + T\Theta'\int_0^\infty \xi_2(x)B_{0,2}(x)dx, \qquad (3.22) \\ h_i &= T(1-f_1)\left(\Theta'\int_0^\infty \xi_1(x)A_{i,1}(x)dx + \Theta\int_0^\infty \xi_1(x)A_{i-1,1}(x)dx\right) \\ &+ T(1-f_2)\left(\Theta'\int_0^\infty \xi_2(x)A_{i,2}(x)dx + \Theta\int_0^\infty \xi_2(x)A_{i-1,2}(x)dx\right) \\ &+ T\left(\Theta'\int_0^\infty \xi_1(x)B_{i,1}(x)dx + \Theta\int_0^\infty \xi_1(x)B_{i-1,1}(x)dx\right) \\ &+ T\left(\Theta'\int_0^\infty \xi_2(x)B_{i,2}(x)dx + \Theta\int_0^\infty \xi_2(x)B_{i-1,2}(x)dx\right), \quad i \geq 1, \qquad (3.23) \end{aligned}$$

where T represents the normalizing constant. To derive the following expression we multiply Equations (3.22)-(3.23) by z^i and then summing over the range of i, we obtain as

$$H(z) = \frac{\begin{array}{l}T\lambda V_0(\Theta z+\Theta')(1-z)\big[((1-f_1)+f_1B_1^*(\lambda-\lambda z))p_1B_1^*(\lambda-\lambda z) \\ \qquad + ((1-f_2)+f_2B_2^*(\lambda-\lambda z))p_2B_2^*(\lambda-\lambda z)\big]\end{array}}{\begin{array}{l}z-(\Theta z+\Theta')\big[((1-f_1)+f_1B_1^*(\lambda-\lambda z))p_1B_1^*(\lambda-\lambda z) \\ \qquad + ((1-f_2)+f_2B_2^*(\lambda-\lambda z))p_2B_2^*(\lambda-\lambda z)\big]\end{array}}. \tag{3.24}$$

Taking limit $z \to 1$ in Equation (3.24), we get

$$T = \frac{\Theta' - p_1\lambda\beta_1^{(1)}(1+f_1) - p_2\lambda\beta_2^{(1)}(1+f_2)}{\lambda V_0}. \tag{3.25}$$

Now using Equation (3.25) in Equation (3.24), we get $H(z)$ as given in Equation (3.21). □

Remark 3.5. If we take $p_1 = 1$ or $p_2 = 0$, that is, there is no STS in the system and $(f_1, f_2) = 0$ means no re-service in the system, then Equation (3.21) reduces to

$$H(z) = \frac{(\Theta'-\rho)(1-z)(\Theta z+\Theta')B_1^*(\lambda-\lambda z)}{[(\Theta z+\Theta')B_1^*(\lambda-\lambda z)-z]}. \tag{3.26}$$

This is the PGF of queue size distribution at departure epoch which matches with the Equation (14) of [27].

Theorem 3.6. *Based on the PGF of the queue size distribution at departure epoch, we can generalize the Pollaczek Khinchine formula.*

Proof. Takagi [26] in the page number (51) states the behavior of queue size distribution at departure epoch of $M/G/1$ queue with Bernoulli feedback system, let $\Phi(z)$ be the PGF of queue size distribution at a departure epoch then by utilizing the argument of [26], we have

$$\Phi(z) = \frac{H(z)}{\Theta z + \Theta'}. \tag{3.27}$$

Now using Equation (3.21) in Equation (3.27), we get

$$\Phi(z) = \frac{\begin{array}{r}(\Theta' - \rho)(z-1)\big[((1-f_1) + f_1 B_1^*(\lambda - \lambda z))p_1 B_1^*(\lambda - \lambda z) \\ + ((1-f_2) + f_2 B_2^*(\lambda - \lambda z))p_2 B_2^*(\lambda - \lambda z)\big]\end{array}}{\begin{array}{r}z - (\Theta z + \Theta')\big[((1-f_1) + f_1 B_1^*(\lambda - \lambda z))p_1 B_1^*(\lambda - \lambda z) \\ + ((1-f_2) + f_2 B_2^*(\lambda - \lambda z))p_2 B_2^*(\lambda - \lambda z)\big]\end{array}}, \tag{3.28}$$

which satisfies the well known Pollaczek Khinchine formula. □

Remark 3.7. In the above result setting $p_1 = 1$, $p_2 = 0$, $(f_1, f_2) = 0$, we get the result which was derived by [25] in the standard $M/G/1$ queueing system with Bernoulli feedback. That is

$$\Phi(z) = \frac{(\Theta' - \rho)(z-1)B_1^*(\lambda - \lambda z)}{z - (\Theta z + \Theta')B_1^*(\lambda - \lambda z)}.$$

Remark 3.8. Let L^+ be the mean queue size at departure epoch. Differentiating Equation (3.28) with respect to z and then setting $z = 1$, we get

$$\begin{aligned} L^+ &= \left(\frac{d\Phi(z)}{dz}\right)_{z=1} = \frac{\rho(1-\rho)}{(\Theta' - \rho)} + \frac{\lambda^2\left(\left(\beta_1^{(2)} + \beta_2^{(2)} f_1\right)p_1 + (1+f_2)\,p_2\beta_2^{(2)}\right)}{2(\Theta' - \rho)} \\ &\quad + \frac{\lambda^2\left(p_1\left(\beta_1^{(1)}\right)^2(f_1 - 2) + p_2\left(\beta_2^{(1)}\right)^2 f_2\right)}{(\Theta' - \rho)}. \end{aligned} \tag{3.29}$$

3.3. LST of response time distribution

An arbitrary customer's response time is the time duration associated from its arrival epoch to the instant of departure. The PDF of the response time is dependable on the service discipline.

Theorem 3.9. *If we denote the response time of an arbitrary customer as R, then subject to the stability condition where* $\frac{\rho}{\Theta'} < 1$, *we can determine the corresponding Laplace-Stieltjes transform (LST) of the response time* $(R*(S))$ *and the expressions for the value is*

$$R^*(s) = \frac{s(\Theta' - \rho)B^*(s)}{(s-\lambda) + (\lambda - s\Theta)B^*(s)}. \tag{3.30}$$

Proof. In order to find $R^*(s)$, we need to obtain some important results. Let β_T denotes the time interval from the instant of starting any type of service to the final departure epoch with no or at least one Bernoulli feedback. Further, assume that the LST of β_T is denoted by $\beta_T^*(s)$. Then $\beta_T^*(s)$ is given by (see [24])

$$\beta_T^*(s) = \sum_{n=0}^{\infty} \Theta' B^*(s)[\Theta B^*(s)]^n = \frac{\Theta' B^*(s)}{1 - \Theta B^*(s)}. \tag{3.31}$$

Now, let D be the delay of a test unit and $D^*(s)$ be the corresponding LST. The delay of the test unit of the first unit in $M^X/G/1$ queueing system is *equivalent* to the delay of a test unit in an ordinary $M/G/1$ queueing system with Bernoulli feedback, see [4]. Therefore, $D^*(s)$ is given by

$$D^*(s) = \frac{s(1-\rho/\Theta')}{s-\lambda\left(1-\beta_T^*(s)\right)}. \tag{3.32}$$

Since, the response time constitutes with two important random variables: β_T and D. Thus, the LST of the response time in our system under study is given by

$$R^*(s) = D^*(s)\beta_T^*(s). \tag{3.33}$$

Using Equations (3.31)-(3.32) in Equation (3.33) gives Equation (3.30).

There are different ways to admit to the system for taking the service again of the unsuccessful unit. In [25], a feedback customer takes service by joining the tail of the queue; in [26] such a customer joins the queue under random order service discipline. In our model, we have considered that the customer joins the tail of the original queue and is immediately taken for service again and again. However, the expected response time performs the same for all the different service disciplines. □

Remark 3.10. Differentiating Equation (3.30) with respect to s and then putting $s=0$ and multiplying the obtained result by -1, we get the mean response time ($E(R)$) of the model as

$$E(R) = \frac{\left(\frac{2}{\lambda}\right)\rho(1-\rho)+\lambda\left(p_1\beta_1^{(2)}+p_2\beta_2^{(2)}\right) - \lambda p_1\left(2(2-f_1)\left(\beta_1^{(1)}\right)^2 - f_1\beta_1^{(2)}\right)+\lambda p_2\left(2\left(\beta_2^{(1)}\right)^2+\beta_2^{(2)}\right)f_2}{2(\Theta'-\rho)}. \tag{3.34}$$

Remark 3.11. Comparing $E(R)$ and L^+ from Equations (3.29) and (3.34), we have $\lambda L^+ = E(R)$ at the departure epoch. This relationship verifies the Little's formula for our mathematical model.

3.4. LST of the inter departure time distribution

The LST of the inter departure time under steady state condition is obtained as

$$T_D^*(s) = \frac{\rho}{\Theta'}\beta_T^*(s)+\left(1-\frac{\rho}{\Theta'}\right)\beta_T^*(s)I_A^*(s), \tag{3.35}$$

where $I_A^*(s)\left(=\frac{\lambda}{\lambda+s}\right)$ is the LST of inter arrival time distribution. Therefore, Equation (3.35) gives

$$T_D^*(s) = \frac{(\rho s+\lambda\Theta')\Big(\left((1-f_1)+f_1B_1^*(s)\right)p_1B_1^*(s)+\left((1-f_2)+f_2B_2^*(s)\right)p_2B_2^*(s)\Big)}{(\lambda+s)\Big(1-\Theta\big(\left((1-f_1)+f_1B_1^*(s)\right)p_1B_1^*(s)+\left((1-f_2)+f_2B_2^*(s)\right)p_2B_2^*(s)\big)\Big)}.$$

3.5. LST of the busy period distribution

The term busy period refers to a specific time duration when the server is consistently engaged and remains occupied until it eventually becomes idle again. If T_{bp} indicates the length of the busy period and $T_{bp}^*(s)$ is the LST of the busy period distribution then we have

$$T_{bp}^*(s) = L_{mst}^*\left(s+\lambda-\lambda T_{bp}^*(s)\right). \tag{3.36}$$

This is written from the standard argument derived by [25]. $L_{mst}^*(s)$ indicates the LST of the modified service time with no or at least one Bernoulli feedback; i.e., here $L_{mst}^*(s) = \beta_T^*(s)$.

Theorem 3.12. *If* $\frac{\rho}{\Theta'} < 1$*, then the first and second moment of the length of the busy period are given as*

$$\begin{aligned}
\textit{first moment} &= \frac{\beta_2^{(1)} + f_2\beta_2^{(1)} + p_1\left((\beta_1^{(1)} - \beta_2^{(1)}) - (f_2\beta_2^{(1)} - f_1\beta_1^{(1)})\right)}{\Theta' - \lambda\left(\beta_2^{(1)} + f_2\beta_2^{(1)} + p_1\left((\beta_1^{(1)} - \beta_2^{(1)}) - (f_2\beta_2^{(1)} - f_1\beta_1^{(1)})\right)\right)}, \\
\textit{second moment} &= \frac{\Theta'^3\Delta(p_1,\Theta,f_1,f_2)}{\left(\Theta' - \lambda\left(\beta_2^{(1)} + f_2\beta_2^{(1)} + p_1\left((\beta_1^{(1)} - \beta_2^{(1)}) - (f_2\beta_2^{(1)} - f_1\beta_1^{(1)})\right)\right)\right)^3},
\end{aligned}$$

where $\Delta(p_1,\Theta,f_1,f_2) = \Big(\frac{1}{(\Theta-1)^2}(\beta_2^{(2)} + \Theta\beta_2^{(2)} + 2\Theta f_2^2\beta_2^{(2)} + 2\Theta p_1^2(-\beta_1^{(1)} + \beta_2^{(1)} - f_1\beta_1^{(1)} + f_2(\beta_2^{(1)}))^2 + f_2(2(1+\Theta)(\beta_2^{(1)})^2 - (\Theta-1)\beta_2^{(2)}) + p_1(\beta_1^{(2)} - \Theta\beta_1^{(2)} + 4\Theta\beta_1^{(1)}\beta_2^{(1)} - \beta_2^{(2)} - 3\Theta\beta_2^{(2)} + 4\Theta f_2\beta_1^{(1)}\beta_2^{(1)} - 2f_2(\beta_2^{(1)})^2 - 6\Theta f_2(\beta_2^{(1)})^2 - f_2\beta_2^{(2)} + \Theta f_2\beta_2^{(2)} - 4\Theta f_2^2\beta_2^{(2)} + f_1(-2(\Theta-1)\beta_1^{(1)}\beta_1^{(1)} + 4\Theta\beta_2^{(1)}\beta_1^{(1)} + \beta_1^{(2)} - \Theta\beta_1^{(2)} + 4\Theta f_2\beta_1^{(1)}\beta_2^{(1)})))\Big)$.

Proof. To find the first moment of the length of the busy period, the reader can obtain it by taking the derivative of $T_{bp}^*(s)$ with respect to s, then substituting s with 0. Finally, multiply the resulting value by -1. This is also represented the average length of the busy period of the system. Here,

$$E(T_{bp}) = -L_{mst}^{*(1)}(0)\left(1 - \lambda T_{bp}^{*(1)}(0)\right).$$

And the second moment of T_{bp} is

$$E(T_{bp}^2) = \left(\frac{d^2}{ds^2}T_{bp}^*(s)\right)_{s=0} = L_{mst}^{*(2)}(0)\left(1 - \lambda T_{bp}^{*(1)}(0)\right)^2 + L_{mst}^{*(1)}(0)\left(-\lambda T_{bp}^{*(2)}(0)\right).$$

□

4. Embedded Markov chain result (PGF of the queue size distribution at departure epoch)

Suppose, t_i represents the time of i^{th} service completion epoch, specifically indicating the epoch when the total service required by a customer ends. We can observe that a sequence denoted as $X_m = N(t_m + 0)$ (where $N(t_m)$ signifies the number of units in the system at the time instant t_m) forms a discrete time Markov chain (DTMC). This DTMC serves as an embedded Markov renewal process of a continuous time Markov process. The sequence $\{X_m; m \geq 0\}$ exhibits homogeneous DTMC and it is owing to the following transition.

$$X_{m+1} = \begin{cases} L_{m+1} - 1, \text{if } X_m = 0 \\ X_m + L_{m+1} - 1, \text{if } X_m > 0, \end{cases} \tag{4.1}$$

where L_m signifies the quantity of units that have entered the system during the m^{th} total service period.

Furthermore, we introduce $\triangle_{i,m}$ matrix which was originally introduced and examined by [1]. This matrix is closely linked to the transition probability matrix (TPM), denoted as $P = (p_{k,n})$.

A finite or infinite stochastic matrix $P = (p_{k,n})$ is called a $\triangle_{i,m}$ matrix, $m \geq i \geq 1$ if $p_{k,n}$ for $k > m$ and $k - n > i$.

When $i = m$, then $\triangle_{i,m}$ matrix reduces to $\triangle_m$ matrix, which in fact is a special case of $\triangle_2$ matrix. Thus, the TPM $P = (p_{k,n})$ associated with the DTMC $\{X_m; m \geq 1\}$ is of the form.

$$\mathrm{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ 0 & P_{21} & P_{22} & \dots \\ 0 & 0 & P_{32} & \dots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Here,

$$P_{kn} = \begin{cases} \sum_{i=1}^{n+1} \sum_{j=1}^{2} p_j \left((1-f_j) d_{j,n-i+1} + f_j \gamma_{j,n-i+1} \right), & \text{if } k = 0, n = 0 \\ \sum\limits_{j=1}^{2} p_j \left((1-f_j) d_{j,m-i+1} + f_j \gamma_{j,n-i+1} \right), & \text{if } k \geq n-1, n \geq 1 \\ 0, & \text{if } k \geq 1, 0 \leq n \leq k-1. \end{cases} \tag{4.2}$$

Let us define,

$$\begin{aligned} d_{j,n} &= \int\limits_{o}^{\infty} \frac{e^{-\lambda x} (\lambda x)^n}{n!} dB_j(x), \\ \gamma_{j,n} &= \int\limits_{o}^{\infty} \frac{e^{-\lambda x} (\lambda x)^n}{n!} dF_j(x), \end{aligned}$$

where $d_{j,n}$ characterizes the probability that 'n' units enter during the j type of service, and $\gamma_{j,n}$ characterizes the probability that 'n' units enter during the j type of repeated service.

Next, we assume that:

$$\frac{\rho}{\Theta'} = \left(\sum_{j=1}^{2} p_j (\rho_{B_j} + f_j \rho_{F_j}) \right) < 1,$$

where $\rho_{B_j} = \lambda \beta_j^{(1)}$, $\rho_{F_j} = \lambda \gamma_j^{(1)}$ to generate that $\{X_m; m \geq 0\}$ is positive recurrent. Consequently, $\frac{\rho}{\Theta'} < 1$ is the necessary and sufficient condition for the existence of steady state condition. Thus, limiting probability $h_n = \lim\limits_{m \to \infty}$ Prob $(X_m = n)$ exist and is positive. then the Kolmogorov equation associated with DTMC $X_m; m \geq 0$ can be written as

$$h_n = \sum_{k=0}^{\infty} h_k P_{kn}; \; n \geq 0.$$

This implies that, for $n \geq 0$, we have

$$h_n = \sum_{m=1}^{n+1} (h_0(1-z) + h_m) \Big[\sum_{j=1}^{2} p_j ((1-r_j) d_{j,n-m+1} + f_j \gamma_{j,n-m+1}) \Big]; n \geq 0. \tag{4.3}$$

Next, let us define the following PGF's for $h_n, d_{j,n}, \gamma_{j,n}, j \in \{1, 2\}$ as

$$H(z) = \sum_{n=0}^{\infty} z^n h_n,$$

$$D_j(z) = \sum_{n=0}^{\infty} z^n d_{j,n},$$

$$\gamma_j(z) = \sum_{n=0}^{\infty} z^n \gamma_{j,n},$$

respectively.

Then from Equation (4.3) we have

$$H(z) = h_0\ (1-z)\left[\begin{array}{l}\sum\limits_{j=1}^{2} p_j\left((1-f_j)D_j(z)+f_j\gamma_j(z)\right)z^{-1}\\ \quad +\left(H(z)-h_0\right)\left(\sum\limits_{j=1}^{2} p_j\left((1-r_j)D_j(z)+f_j\gamma_j(z)\right)z^{-1}\right)\end{array}\right]. \tag{4.4}$$

Now, because of the presence of convolution, Equation (4.4) can be transformed with the help of PGF's:

$$D_j(z) = B_j^*(\lambda-\lambda z),$$
$$\gamma_j(z) = D_j(z)B_j^*(\lambda-\lambda z),\ \text{for } j=1,2.$$

Therefore, from Equation (4.4), we have

$$H(z) = \frac{h_0[1-\alpha(z)]\left[\sum_{j=1}^{2} p_j((1-f_j)+f_jR_j^*(\lambda-\lambda z))B_j^*(\lambda-\lambda z))\right](\Theta'+\Theta z)}{\left[\sum_{j=1}^{2} p_j((1-f_j)+f_jB_j^*(\lambda-\lambda z))B_j^*(\lambda-\lambda z))(\Theta'+\Theta z)-z\right]}. \tag{4.5}$$

Now, since $\sum\limits_{n=0}^{\infty} h_n = H(1) = 1$, Equation (4.5) yields

$$h_0 = (\Theta'-\rho). \tag{4.6}$$

By observing Equation (4.6) as provided above, it is apparent that $\frac{\rho}{\Theta'} < 1$. This condition serves as the necessary and sufficient requirement for the existence of a steady-state solution in our model. Consequently, with $\frac{\rho}{\Theta'} < 1$ we can derive as follows:

$$(\Theta'+\Theta z)\sum_{j=1}^{2} p_j((1-f_j)+f_jB_j^*(\lambda-\lambda z))B_j^*(\lambda-\lambda z)-z=0,$$

which never vanishes inside the region $|z|\leq 1$, by virtue of Rouche's theorem.

Therefore, by utilizing Equation (4.6) in Equation (4.5) we ultimately obtain the probability generating function (PGF) for the queue size distribution at departure epoch as shown below:

$$H(z) = \frac{(\Theta'-\rho)(1-z)\left[\sum_{j=1}^{2} p_j((1-f_j)+f_jB_j^*(\lambda-\lambda z))B_j^*(\lambda-\lambda z))\right](\Theta'+\Theta z)}{\left[\sum_{j=1}^{2} p_j((1-f_j)+f_jB_j^*(\lambda-\lambda z))B_j^*(\lambda-\lambda z))(\Theta'+\Theta z)-z\right]}, \tag{4.7}$$

where,

$$\rho = p_1\lambda(1+f_1)\beta_1^{(1)} + p_2\lambda(1+f_2)\beta_2^{(1)}.$$

5. Cost analysis

Cost analysis is the most important fact in any practical situation at every stage. It is quite natural that the management of the system desires to minimize the total average cost. In this Section, the cost model for the proposed queueing system is developed and the total expected cost per unit of time is given by

$$TC = C_hM_q + C_0\left(\frac{\rho}{\Theta'}\right) + C_S\left(\frac{1}{\mu_{bc}}\right)$$

where,
C_h is the holding costs per unit time for each customer present in the system,
C_0 is the cost per unit time for keeping the server on and in operations,
C_S is the setup cost per busy cycle.

M_q is the average queue size under steady state which is available in Theorem 3.1, where

$$M_q = \frac{\begin{array}{c} 2f_1p_1\lambda^2\beta_1^{(2)} + (1-f_1)p_1\left[\lambda^2\beta_1^{(2)} + \lambda\beta_1^{(1)}\right] \\ + 2f_2p_2\lambda^2\beta_2^{(2)} + (1-f_2)p_2\left[\lambda^2\beta_2^{(2)} + \lambda\beta_2^{(1)}\right] \end{array}}{2\left[1 - \Theta - p_1\lambda\beta_1^{(1)}(1+f_1) - p_2\lambda\beta_2^{(1)}(1+f_2)\right]}.$$

$\left(\frac{\rho}{\Theta'}\right)$ is the stability condition necessary for the steady state solution exists, where

$$\rho = p_1\lambda(1+f_1)\beta_1^{(1)} + p_2\lambda(1+f_2)\beta_2^{(1)}.$$

And, μ_{bc} is the mean busy cycle, i.e.,

$$\text{Mean busy cycle } (\mu_{bc}) = \text{Mean busy period } (\mu_{bp}) + \text{Mean idle period } (\mu_{ip}),$$

where,

$$\text{Mean busy period } (\mu_{bp}) = \frac{\beta_2^{(1)} + f_2\beta_2^{(1)} + p_1\left((\beta_1^{(1)} - \beta_2^{(1)}) - (f_2\beta_2^{(1)} - f_1\beta_1^{(1)})\right)}{\Theta' - \lambda\left(\beta_2^{(1)} + f_2\beta_2^{(1)} + p_1\left((\beta_1^{(1)} - \beta_2^{(1)}) - (f_2\beta_2^{(1)} - f_1\beta_1^{(1)})\right)\right)},$$

$$\text{Mean idle period}(\mu_{ip}) = \frac{1}{\lambda}.$$

Moreover, we can examine numerically the behavior of the expected cost function under different values of the parameters of the model under study. The default values of the different cost elements are considered as: $C_h = 10, C_0 = 150, C_S = 1000$ and giving the suitable values to the other following parameters satisfies the stability condition. The numerical results displayed in Tables 1 - 3 are obtained for the fixed parameters: $p_1 = 0.6$, $p_2 = 0.4$ and considering that the service time (B_j) follows exponential distribution with parameter $\mu_j, j = 1, 2$. The obtained results in Tables 1 - 3 show the effects of the parameters of the system on the total cost (TC).

From Tables 1 - 3, we notice that for fixed arrival rate (λ), μ_1, μ_2 and Θ, when probability of repeated service increases then TC decreases. This happens because when the same customer remain in his position in queue and take more time for completion of his/her service, then some of the customer which are staying in the queue may departs from the system (impatient due to balking).

Table 1. Effects of different parametric values on the total expected cost (TC) when $\mu_1 = \mu_2 = 5$.

$\mu_1 = \mu_2 = 5$									
	$\Theta = 0.4$			$\Theta = 0.5$			$\Theta = 0.6$		
λ	$f_1 = 0.2$ $f_2 = 0.4$	$f_1 = 0.3$ $f_2 = 0.5$	$f_1 = 0.4$ $f_2 = 0.6$	$f_1 = 0.2$ $f_2 = 0.4$	$f_1 = 0.3$ $f_2 = 0.5$	$f_1 = 0.4$ $f_2 = 0.6$	$f_1 = 0.2$ $f_2 = 0.4$	$f_1 = 0.3$ $f_2 = 0.5$	$f_1 = 0.4$ $f_2 = 0.6$
0.2	196.03	195.67	195.30	195.25	194.81	194.37	194.07	193.52	192.98
0.4	358.08	354.69	351.30	349.73	345.66	341.59	337.24	332.16	327.08
0.6	486.18	477.11	468.04	463.54	452.67	441.79	429.70	416.14	402.56
0.8	580.42	563.04	545.66	536.82	516.00	495.18	471.83	445.94	420.07
0.9	614.88	592.36	569.84	558.34	531.40	504.46	474.27	440.88	407.59
1.0	640.92	612.62	584.33	569.85	536.03	502.25	464.56	422.95	381.69

Also, from Tables 1 - 3, we observe that for fixed μ_1, μ_2, f_1, f_2 and λ values when the probability of Bernoulli feedback Θ increases then, the expected total cost (TC) decreases. Because when after taking the service if the dis-satisfied customer increases then obviously some of the customer in the queue which are waiting for taking service may departs from the system and therefore TC decreases, which generates a negative impact in the economy.

Table 2. Effects of different parametric values on the total expected cost (TC) when $\mu_1 = \mu_2 = 10$.

$\mu_1 = \mu_2 = 10$									
	$\Theta = 0.4$			$\Theta = 0.5$			$\Theta = 0.6$		
λ	$f_1 = 0.2$ $f_2 = 0.4$	$f_1 = 0.3$ $f_2 = 0.5$	$f_1 = 0.4$ $f_2 = 0.6$	$f_1 = 0.2$ $f_2 = 0.4$	$f_1 = 0.3$ $f_2 = 0.5$	$f_1 = 0.4$ $f_2 = 0.6$	$f_1 = 0.2$ $f_2 = 0.4$	$f_1 = 0.3$ $f_2 = 0.5$	$f_1 = 0.4$ $f_2 = 0.6$
0.2	198.00	197.82	197.64	197.60	197.38	197.16	197.01	196.73	196.46
0.4	378.97	377.27	375.57	374.77	372.73	370.69	368.47	365.92	363.38
0.6	542.90	538.36	533.81	531.50	526.05	520.59	514.41	507.59	500.77
0.8	689.81	681.09	672.36	667.81	657.34	646.87	634.84	621.76	608.68
0.9	756.88	745.57	734.26	728.31	714.74	701.17	685.50	668.54	651.58
1.0	819.70	805.47	791.24	783.71	766.64	749.56	729.79	708.46	687.12

Table 3. Effects of different parametric values on the total expected cost (TC) when $\mu_1 = \mu_2 = 20$.

$\mu_1 = \mu_2 = 20$									
	$\Theta = 0.4$			$\Theta = 0.5$			$\Theta = 0.6$		
λ	$f_1 = 0.2$ $f_2 = 0.4$	$f_1 = 0.3$ $f_2 = 0.5$	$f_1 = 0.4$ $f_2 = 0.6$	$f_1 = 0.2$ $f_2 = 0.4$	$f_1 = 0.3$ $f_2 = 0.5$	$f_1 = 0.4$ $f_2 = 0.6$	$f_1 = 0.2$ $f_2 = 0.4$	$f_1 = 0.3$ $f_2 = 0.5$	$f_1 = 0.4$ $f_2 = 0.6$
0.2	198.10	198.91	198.81	198.80	198.69	198.58	198.50	198.36	198.22
0.4	389.47	388.62	387.77	387.36	386.34	385.32	384.21	382.93	381.66
0.6	571.41	569.14	566.86	565.70	562.97	560.24	557.13	553.72	550.31
0.8	744.83	740.47	736.10	733.81	728.57	723.33	717.27	710.72	704.18
0.9	828.35	822.69	817.03	814.03	807.23	800.44	792.55	784.06	775.57
1.0	909.73	902.61	895.48	891.69	883.14	874.59	864.63	853.95	843.26

Upon examining from Tables 1 - 3 we observe that as the values of μ_1 and μ_2 increase, the total expected cost (TC) also increases if the other parametric values are fixed.

On the other hand, from Table 1 we notice that the minimum total incurred cost is 192.98 for the optimal parameters ($\Theta = 0.6, f_1 = 0.4, f_2 = 0.6, \mu_1 = \mu_2 = 5, \lambda = 0.2$), from Table 2 we observe that the minimum total incurred cost is 196.46 for the optimal parameters ($\Theta = 0.6, f_1 = 0.4, f_2 = 0.6, \mu_1 = \mu_2 = 10, \lambda = 0.2$), and from Table 3 we notice that the minimum total incurred cost is 198.22 for the optimal parameters ($\Theta = 0.6, f_1 = 0.4, f_2 = 0.6, \mu_1 = \mu_2 = 20, \lambda = 0.2$).

Therefore, from the above numerical analysis we observed that the influence of parameters on the total cost (TC) in the system coincides with the practical situations.

6. Simulation study

In this Section, we show how different parameters affect the system performance by means of some graphical results. In Figure 2 - Figure 5, we assume that the service time distribution follows exponential distribution with parameter $\mu_j, j = 1, 2$ and we consider the following four cases:

case 1: $\mu_1 = \mu_2 = 3$,
case 2: $\mu_1 = \mu_2 = 10$,
case 3: $\mu_1 = 3 < \mu_2 = 10$,
case 4: $\mu_1 = 10 > \mu_2 = 3$.

The other parametric values for Figure 2 are $p_1 = 0.2, p_2 = 0.8, f_1 = 0.1, f_2 = 0.2, \Theta = 0.3$. We notice from Figure 2 that for each case, the average queue size under steady state is seen to be increased as the customer's arrival rate increases. But in the lower arrival rate mean queue size under steady state is insensitive to each case. For fix and higher arrival rate the mean queue size under steady state is maximum as the condition of case 1 is noticed and minimum as the condition of case 2 is fulfilled. We have also noticed from Figure 2 that in case 4 for fixed and higher arrival rate the mean queue size under steady state is higher as compared to case 3.

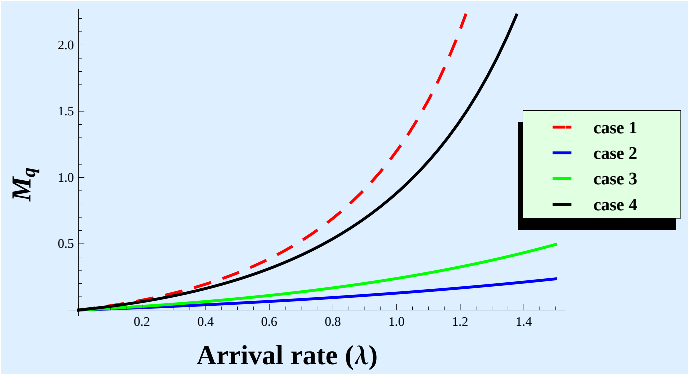

Figure 2. Impact of λ on M_q for different service rates.

The parametric values of Figure 3 are $\lambda = 0.3, f_1 = 0.1, f_2 = 0.2, \Theta = 0.3$ and observe that for fixed and lower p_1 the average queue size under steady state is maximum as the condition of case 1 is fulfilled and minimum as the condition of case 2 is noticed. So we may conclude that for equal and lower μ_1, μ_2 value gives maximum M_q and equal and higher μ_1, μ_2 value gives minimum M_q for increasing value of p_1. In case 3 the average queue size under steady state increases as p_1 increases, in case 4 M_q behaves downward for increasing value of p_1.

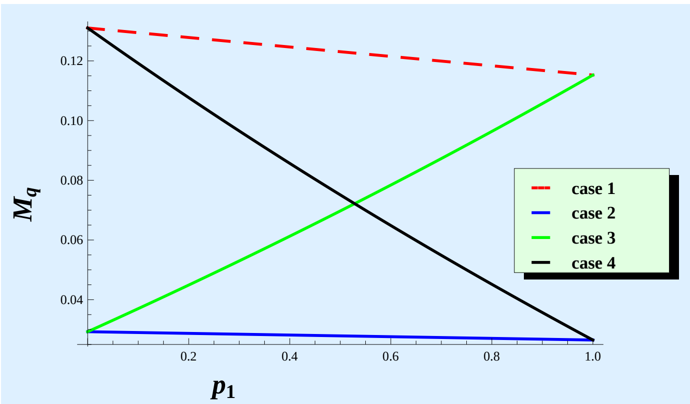

Figure 3. Impact of p_1 on M_q for different service rates.

In Figure 4 the parametric values are $p_1 = 0.2, p_2 = 0.8, f_1 = 0.1, f_2 = 0.2, \Theta = 0.3$ and notice that for all the four cases the effect of arrival rate on W_q is similar as that of M_q in Figure 2. Moreover, in Figure 5 the parametric values are $\lambda = 0.4, f_1 = 0.6, f_2 = 0.8, \Theta = 0.6$ and observe that the effect of p_1 on W_q is similar as that of M_q in Figure 3.

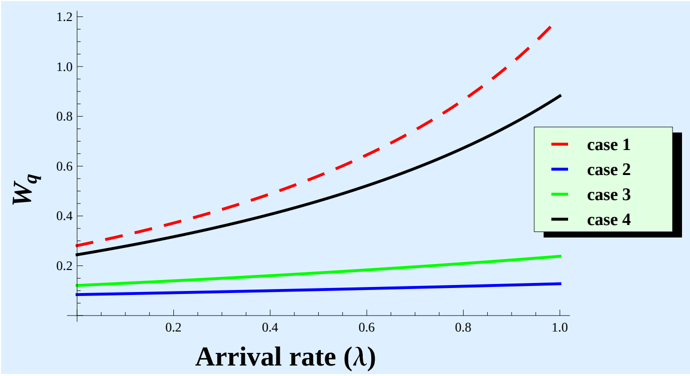

Figure 4. Impact of λ on W_q for different service rates.

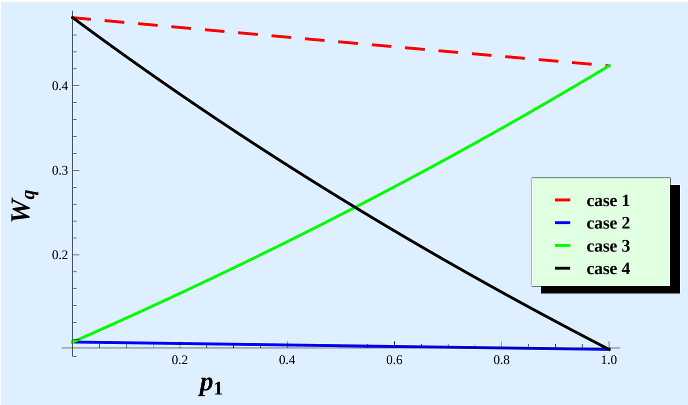

Figure 5. Impact of p_1 on W_q for different service rates.

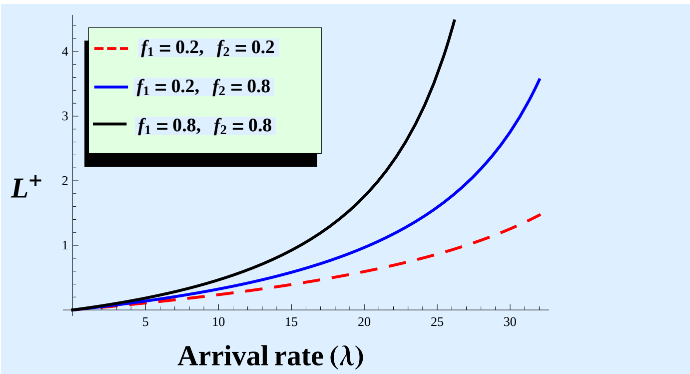

Figure 6. Effects of λ on L^+ for different f_1 and f_2.

In Figure 6, the general service time and repeated service time follow exponential distribution as $B_j \sim Exp(75)$ and $F_j \sim Exp(50)$, respectively, and the other parametric values are $p_1 = 0.6, p_2 = 0.4, \Theta = 0.2$. We observe the mean queue size at departure epoch (L^+) with customer's arrival rate for considering the three situations of repeated probability $f_1 = 0.2, f_2 = 0.2; f_1 = 0.2, f_2 = 0.8$; and $f_1 = 0.8, f_2 = 0.8$. From Figure 6 we have noticed that for all these above mentioned three situations the mean queue size at departure epoch is increases as arrival rate increases, but L^+ is insensitive to the above three situations as the arrival rate is lower. For fixed and higher λ, L^+ is maximum when $f_1 = 0.8, f_2 = 0.8$, but L^+ is minimum when $f_1 = 0.2, f_2 = 0.2$ and when $f_1 = 0.2, f_2 = 0.8$, L^+ seems moderate as compared to other two situations.

Table 4 shows the effect of general service and re-service on the system state probabilities. The general service time follows exponential distribution with parameter μ_j,i.e. $B_j \sim Exp(\mu_j)$ so $\beta_j^{(1)} = \frac{1}{\mu_j}, \beta_j^{(2)} = \frac{2}{\mu_j^2}, j = 1, 2$ and consider the two situations: $\mu_1 = \mu_2 = 20$ and $\mu_1 = 30, \mu_2 = 35$.

Table 4. Effect of different parameters on the system state probabilities.

$\lambda = 5, \Theta = 0.4, \mu_1 = \mu_2 = 20$							
p_1	P_{B_1}	P_{B_2}	f_1	f_2	ρ	P_{F_1}	P_{F_2}
0.0000	0.0000	0.4167	0.0100	1.0000	0.5000	0.0000	0.4167
0.1000	0.0417	0.3750	0.1100	0.9000	0.4553	0.0046	0.3375
0.3000	0.1250	0.2917	0.3100	0.7000	0.3957	0.0387	0.2042
0.4000	0.1667	0.2500	0.4100	0.6000	0.3810	0.0683	0.1500
0.5000	0.2083	0.2083	0.5100	0.5000	0.3762	0.1062	0.1042
0.6000	0.2500	0.1667	0.6100	0.4000	0.3815	0.1525	0.0667
0.7000	0.2917	0.1250	0.7100	0.3000	0.3967	0.2071	0.0375

$\lambda = 5, \Theta = 0.4, \mu_1 = 30, \mu_2 = 35$							
p_1	P_{B_1}	P_{B_2}	f_1	f_2	ρ	P_{F_1}	P_{F_2}
0.0000	0.0000	0.2381	0.0100	1.0000	0.2857	0.0000	0.2381
0.1000	0.0278	0.2143	0.1100	0.9000	0.2628	0.0030	0.1928
0.2000	0.0555	0.1905	0.2100	0.8000	0.2460	0.0117	0.1524
0.3000	0.0833	0.1667	0.3100	0.7000	0.2355	0.0258	0.1167
0.4000	0.1111	0.1428	0.4100	0.6000	0.2311	0.0455	0.0857
0.5000	0.1389	0.1190	0.5100	0.5000	0.2330	0.0708	0.0595
0.6000	0.1667	0.0952	0.6100	0.4000	0.2410	0.1017	0.0381
0.7000	0.1944	0.0714	0.7100	0.3000	0.2552	0.1380	0.0214

One can observe from Table 4 that for $\mu_1 = \mu_2 = 20$, P_{B_1} is increasing and P_{B_2} is decreasing with p_1 since P_{B_1} and P_{B_2} are proportional to p_1 and $p_2 = (1-p_1)$, respectively. Further, since P_{F_1} is proportional to p_1 and f_1, thus P_{F_1} is increasing with p_1 and f_1. Moreover, as P_{F_2} is proportional to $p_2 = (1 - p_1)$ and f_2, thus P_{F_2} is decreasing with the increment in p_1 and decrement in f_2. Now for $\mu_1 = 30, \mu_2 = 35$ similar behavior of P_{B_1}, P_{B_2}, P_{F_1}, and P_{F_2} are observed. However, the probabilities are lower as compared to $\mu_1 = \mu_2 = 20$.

Table 5. Effect of arrival rate and feedback probability on mean response time

$\rho \downarrow$	$\lambda \downarrow$	$\Theta \rightarrow$ 0.2667	0.3556	0.4444	0.5333	0.6222	0.7111	0.8000
0.0190	0.0500	0.5194	0.5932	0.6915	0.8287	1.0341	1.3746	2.0497
0.0359	0.0944	0.5205	0.5965	0.6985	0.8427	1.0617	1.4348	2.2119
0.0528	0.1389	0.5216	0.6000	0.7061	0.8577	1.0923	1.5036	2.4114
0.0697	0.1833	0.5229	0.6037	0.7142	0.8741	1.1262	1.5829	2.6625
0.0866	0.2278	0.5241	0.6077	0.7228	0.8919	1.1641	1.6755	2.9883
0.1034	0.2722	0.5255	0.6119	0.7321	0.9113	1.2066	1.7850	3.4282
0.1203	0.3167	0.5269	0.6163	0.7422	0.9327	1.2547	1.9163	4.0546
0.1372	0.3611	0.5286	0.6211	0.7530	0.9562	1.3095	2.0770	5.0180
0.1541	0.4056	0.5301	0.6261	0.7648	0.9823	1.3727	2.2780	6.6905
0.1710	0.4500	0.5317	0.6316	0.7776	1.0113	1.4461	2.5365	10.3110

Table 5 shows the variation of the mean response time with arrival rate and feedback probability for exponentially distributed service time with parameter $\mu_j, j = 1, 2$. One can observe from the Table 5 that for fixed ρ and λ, the mean response time increases with the increment in the value of Θ. Further, for any fixed value of Θ, the mean response time also increases with the increment in ρ and λ. This variation seems prevalent with the increasing value of Θ and λ, the numbers of customers and feedback customers become larger as such the mean response time increases rapidly. Further, the effect of re-service on the mean response time is shown in Table 6.

Table 6. Effect of re-service on the mean response time.

		$\mu_1 = \mu_2 = 3$			$\mu_1 = \mu_2 = 10$		
	Θ	$f_1 = 0.1$ $f_2 = 0.3$	$f_1 = 0.6$ $f_2 = 0.2$	$f_1 = 0.8$ $f_2 = 1.0$	$f_1 = 0.1$ $f_2 = 0.3$	$f_1 = 0.6$ $f_2 = 0.2$	$f_1 = 0.8$ $f_2 = 1.0$
$\lambda = 0.15$ $p_1 = 0.2$	0.00000	0.43992	0.44732	0.69990	0.12771	0.12977	0.19990
	0.08333	0.48287	0.49104	0.77114	0.13957	0.14182	0.21867
	0.16667	0.53510	0.54423	0.85853	0.15385	0.15634	0.24135
	0.25000	0.60001	0.61034	0.96826	0.17138	0.17417	0.26924
	0.33333	0.68284	0.69473	1.11015	0.19343	0.19659	0.30445
	0.41667	0.79220	0.80621	1.30077	0.22199	0.22562	0.35025
	0.50000	0.94326	0.96031	1.57041	0.26044	0.26473	0.41228
	0.58333	1.16552	1.18722	1.98109	0.31500	0.32023	0.50099
	0.66667	1.52481	1.55455	2.68261	0.39849	0.40519	0.63836
	0.75000	2.20431	2.25104	4.1533	0.54218	0.55148	0.87950

		$\mu_1 = 3, \mu_2 = 10$			$\mu_1 = 10, \mu_2 = 3$		
	Θ	$f_1 = 0.1$ $f_2 = 0.3$	$f_1 = 0.6$ $f_2 = 0.2$	$f_1 = 0.8$ $f_2 = 1.0$	$f_1 = 0.1$ $f_2 = 0.3$	$f_1 = 0.6$ $f_2 = 0.2$	$f_1 = 0.8$ $f_2 = 1.0$
$\lambda = 0.15$ $p_1 = 0.2$	0.00000	0.17657	0.20509	0.28585	0.39099	0.37177	0.61327
	0.08333	0.19310	0.22437	0.31308	0.42882	0.40763	0.67474
	0.16667	0.21304	0.24766	0.34604	0.47475	0.45115	0.74992
	0.25000	0.23759	0.27634	0.38678	0.53170	0.50508	0.84395
	0.33333	0.26852	0.31253	0.43838	0.60418	0.57364	0.96495
	0.41667	0.30871	0.35963	0.50586	0.69953	0.66374	1.12644
	0.50000	0.36305	0.42345	0.59790	0.83061	0.78744	1.35285
	0.58333	0.44062	0.51481	0.73089	1.02215	0.96778	1.69317
	0.66667	0.56032	0.65642	0.93995	1.32852	1.25526	2.26227
	0.75000	0.76934	0.90552	1.31654	1.89714	1.78571	3.40761

Now, we present some numerical computations to show the effect of different parameters on the mean response time, see Figures 7 - 10.

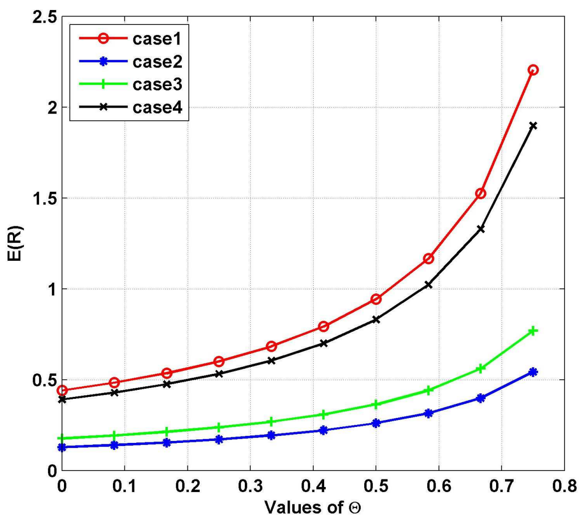

Figure 7. $\lambda = 0.15, p_1 = 0.2, f_1 = 0.1, f_2 = 0.3$.

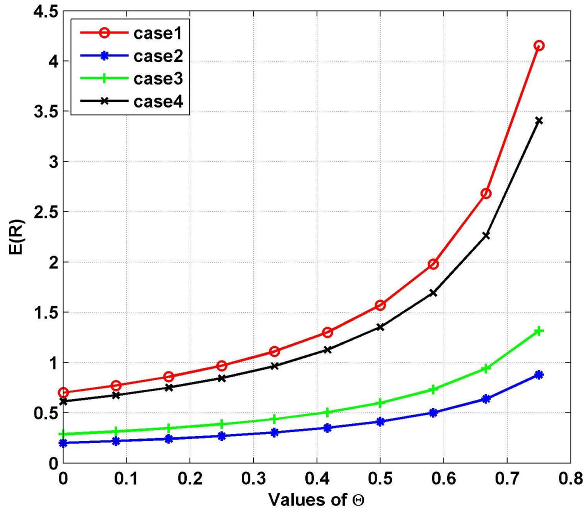

Figure 8. $\lambda = 0.15, p_1 = 0.2, f_1 = 0.8, f_2 = 1.0.$

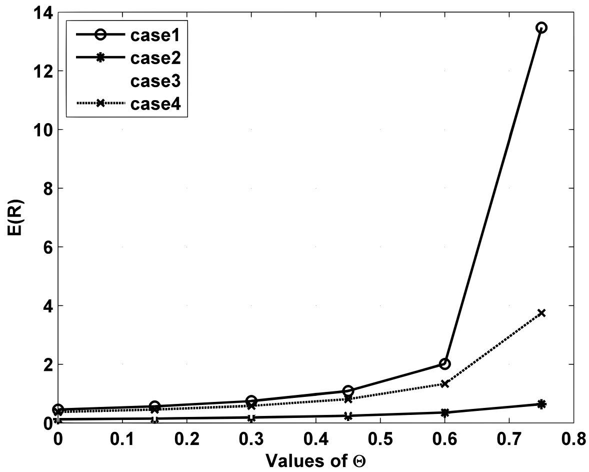

Figure 9. $\lambda = 0.55, p_1 = 0.4, f_1 = 0.1, f_2 = 0.3.$

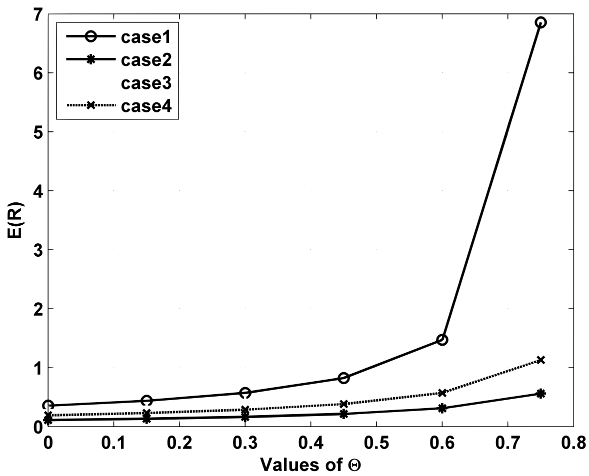

Figure 10. $\lambda = 0.55, p_1 = 0.8, f_1 = 0.1, f_2 = 0.3.$

For this, we have chosen for Figures 7-10 that the service time distribution follows the exponential distribution with parameter $\mu_j,\ j = 1, 2$. For the numerical observation, we have also considered the following cases under $\frac{\rho}{\Theta'} < 1$, that is,

case 1: $\mu_1 = \mu_2 = 3$,
case 2: $\mu_1 = \mu_2 = 10$,
case 3: $\mu_1 = 3 < \mu_2 = 10$,
case 4: $\mu_1 = 10 > \mu_2 = 3$.

and observe from Figures 7 - Figure 10 that the mean response time is highest for case 1 and lowest for case 2, under the condition that the other parameters are fixed. Therefore it is clear that in the situation of $\mu_1 = \mu_2$ the mean response time decreases when μ_i's are increases.

In Figure 11, we consider the parametric values as $p_1 = 0.2$, $f_1 = 0.6$, $p_2 = 0.8$, $f_2 = 0.4$, $\rho = 0.2$, $\Theta = 0.2$, $\mu_1 = 5$, $\mu_2 = 7$ to exhibit the behavior of mean response time of an arbitrary customer on the arrival rate for different service time distributions: hyper-exponential, exponential and 2-stage Erlang distribution with parameter μ_j, $j = 1, 2$. From the Figure 11, we see that for higher arrival rate, the mean response time is larger when service time follows hyper-exponential distribution as compared to 2-stage Erlang and exponentially distributed service time.

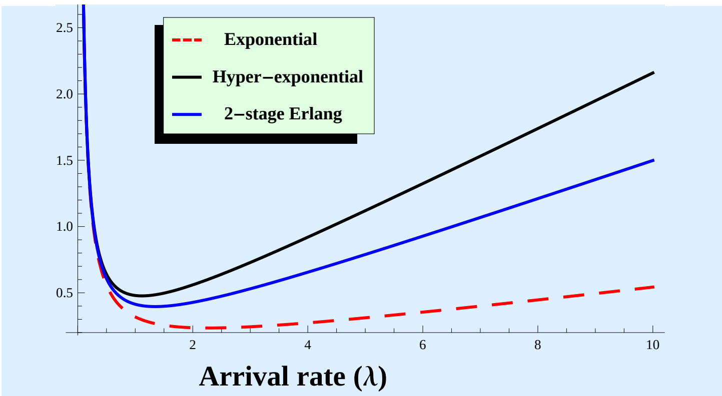

Figure 11. Effects of λ on $E(R)$.

From Figures 12 - 15 we consider that the service time distribution follows exponential, hyper-exponential and 2-stage Erlang with parameter $\mu_j, j = 1, 2$.

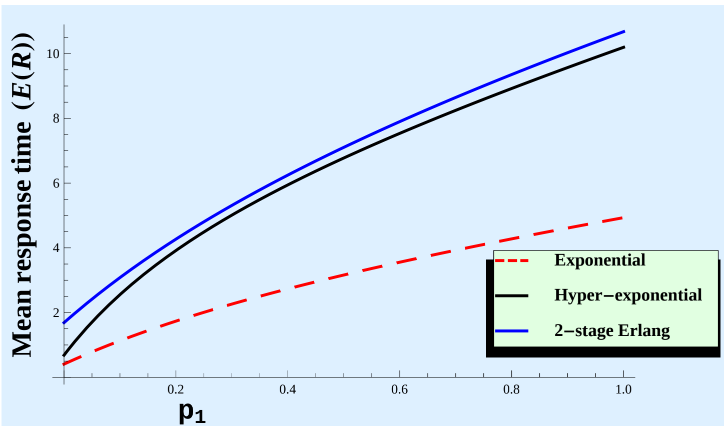

Figure 12. Effects of p_1 on $E(R)$.

The other parameters for Figure 12 are $\lambda = 3, f_1 = 0.6, f_2 = 0.4, \Theta = 0.2, \mu_1 = 0.8, \mu_2 = 2.0$ and notice that mean response time increases as p_1 increases. For fixed p_1, mean response time is higher when the service time distribution follows 2-stage Erlang whereas it provides the lower mean response time for exponential distribution. The customers' mean response time for hyper-exponential service time distribution stays in between exponential and 2-stage Erlang service time distribution. Therefore, we may conclude that as the service time follows exponential distribution then it is favorable for a customer.

In Figure 13, presuming the model parameters as $\lambda = 3, f_1 = 0.5, f_2 = 0.4, \Theta = 0.2, \mu_1 = 0.4, \mu_2 = 0.8$ and we observe that mean response time is higher as p_2 is lower. For the three different service time distributions it is noticed that as the value of p_2 increasing the mean response time is decreasing and like the Figure 12, Figure 13 also shows that for the exponential service time distribution the mean response time is minimum as compared to the other two distributions.

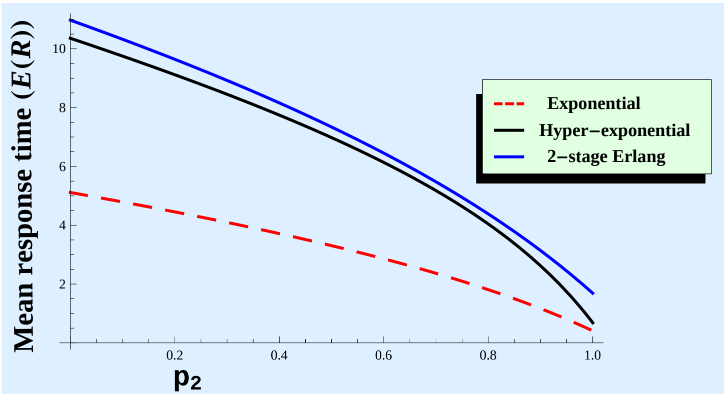

Figure 13. Effects of p_2 on $E(R)$.

In Figure 14, considering the model parameters as $\lambda = 3, p_1 = 0.6, f_2 = 0.4, \Theta = 0.2, \mu_1 = 0.4, \mu_2 = 0.8$ and we have noticed that as f_1 increases mean response time behaves in opposite direction. For the exponential service time distribution mean response time is more less by comparing when the service time distribution follows hyper-exponential or 2-stage Erlang distribution. For fixed value of f_1, mean response time goes to maximum when the service time distribution is 2-stage Erlang.

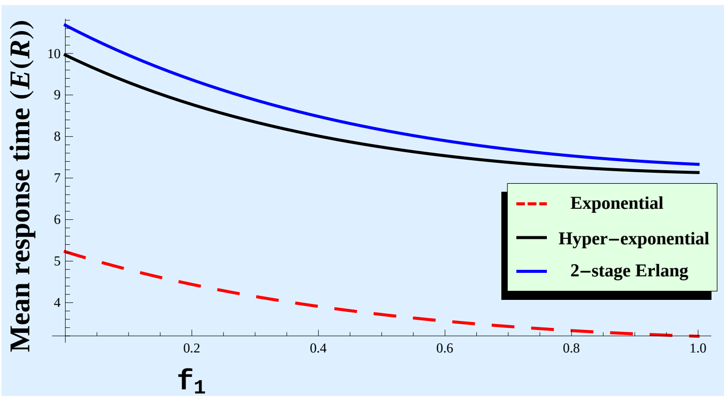

Figure 14. Effects of f_1 on $E(R)$.

The parameters for Figure 15 as $\lambda = 3, p_1 = 0.6, f_1 = 0.4, \Theta = 0.2, \mu_1 = 0.8, \mu_2 = 1.5$ and we have seen that for each service time distribution mean response time is lower when f_2 is higher. For fixed f_2, mean response time is the lowest when service time distribution follows exponential distribution.

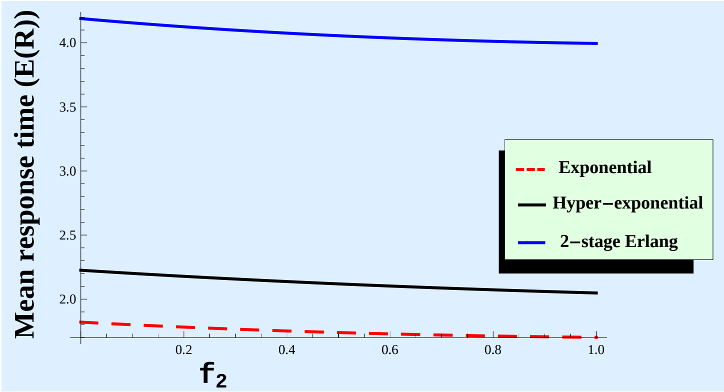

Figure 15. Effects of f_2 on $E(R)$.

7. Application

The considered queueing model has applications in the areas of manufacturing, production, telecommunication systems, etc. A detailed application of the considered model in digital communication network system is presented below.

The electronic mail system, commonly known as email, is a method of sending and receiving digital messages over the internet or a computer network. It has revolutionized communication by providing a fast, efficient, and convenient way to exchange information. It operates on a set of protocols that define how email messages are composed, transmitted, and delivered. The email system operates based on specific protocols, including SMTP (Simple Mail Transfer Protocol) for sending messages and either POP (Post Office Protocol) or IMAP (Internet Message Access Protocol) for receiving and accessing emails. These protocols guarantee the accurate transmission and delivery of emails to their intended recipients.

SMTP is a standard protocol for the transmission of email via the internet. It (SMTP) operates on top of the Transmission Control Protocol (TCP), which is one of the core protocols of the internet protocol suite. TCP provides reliable, connection-oriented communication between two devices and it ensures that data sent over the network arrives in the correct order and without errors. When an email client or server wants to send an email message using SMTP, it establishes a TCP connection with the destination mail server. The SMTP client then initiates a conversation with the server, following a specific set of commands and responses defined in the SMTP protocol. It's worth noting that while port 25 is the standard port for SMTP, some email service providers use alternative ports (e.g., 587) to enhance security and prevent spam abuse. Overall, SMTP over TCP provides a robust and widely adopted mechanism for sending and receiving email messages on the internet.

The Post Office Protocol (POP) is an application-layer protocol used for retrieving email messages from a remote mail server. There are two main versions of POP: POP3 and POP2. POP3 (Post Office Protocol version 3) is the most widely used version. It allows email clients to download messages from the mail server to a local device (such as a computer or a mobile phone) for offline access. The protocol typically operates over TCP/IP (Transmission Control Protocol/Internet Protocol). The POP client and POP server communicate with each other using TCP connections. TCP guarantees that the data exchanged between the client and server is delivered accurately and in the correct sequence. However, IMAP (Internet Message Access Protocol) is a protocol used by email clients to retrieve and manage email messages from a mail server. IMAP operates over TCP/IP, which provides the underlying network communication. When an email client connects to an IMAP server, it establishes a TCP connection to the server's IMAP port (usually port 143 for non-encrypted connections or port 993 for encrypted connections using Secure Sockets Layer (SSL) or Transport Layer Security (TLS)). The client then communicates with the server using the IMAP commands and responses over the established TCP connection.

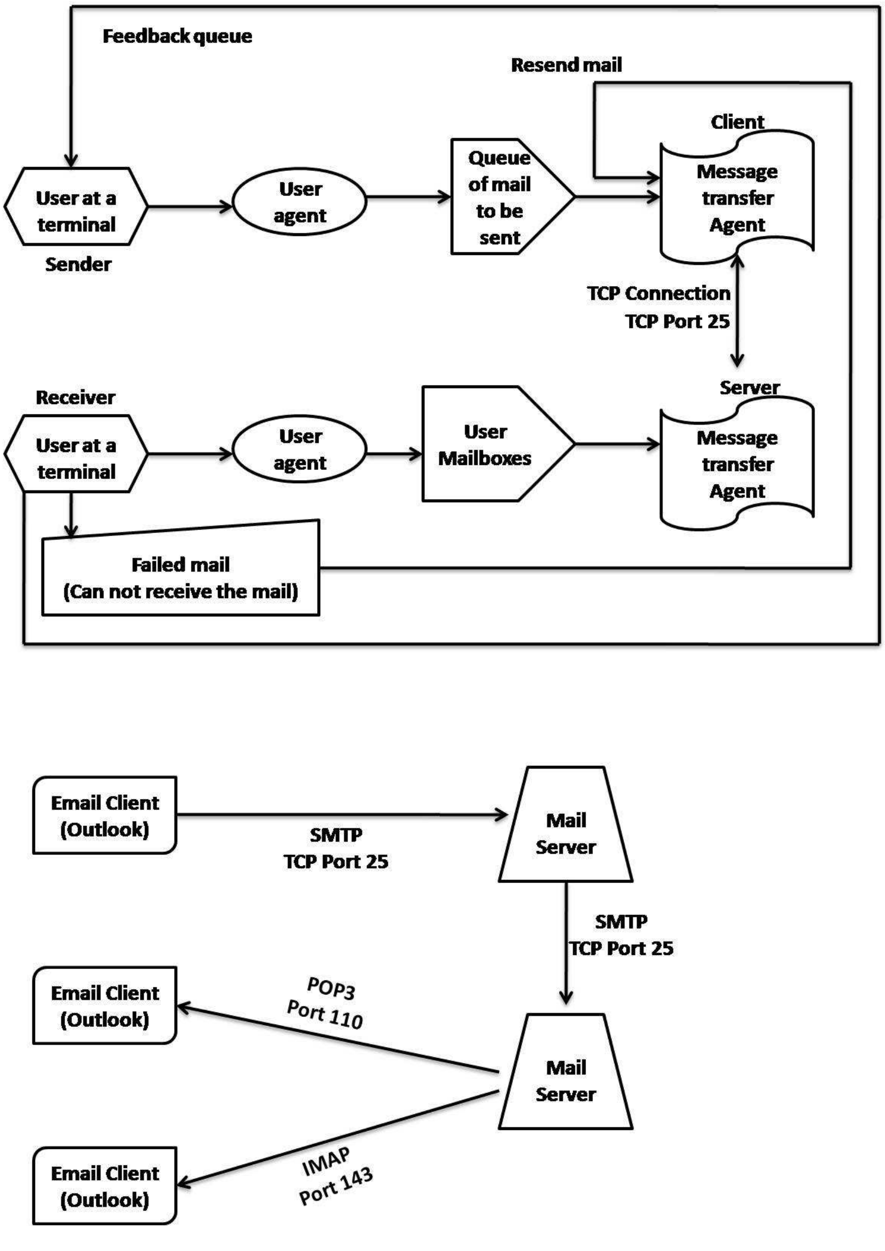

Figure 16. Simple mail transfer protocol (SMTP) model.

Typically, the message handling request comes to the SMTP server follows Poisson process, and the message may select one of the two types of service (POP or IMAP) at the SMTP server (depending upon the choice of the sender). The message transfer agent (MTA) may resend the message again because of the failure of the previous message. For example, since, in an end-to-end application related to SMTP (Simple Mail Transfer Protocol), it is important for the server to be available whenever a client transmits mail. Since, SMTP is the standard protocol used for sending email messages over the internet, and when a client wants to send an email, it connects to an SMTP server and submits the message for delivery. The server then processes the email, performs necessary checks and validations, and attempts to deliver it to the recipient's email server. If the server is not available or offline, the client will not be able to transmit the mail, and the delivery process will fail.

It's crucial for the SMTP server to be consistently available to ensure reliable email transmission. Organizations and email service providers typically maintain redundant and highly available server infrastructure to minimize downtime and provide uninterrupted email services. This ensures that clients can reliably transmit their mail at any time, and the server can handle the delivery process efficiently. On the other hand, if the recipient's mail server receives an email but is not satisfied with it due to late arrival, there are a few potential scenarios and actions that can be taken, say, bouncing the email, placing the email in the spam folder, etc. The recipient's mail server might reject the email and send a bounce message to the sender indicating that the email was not delivered due to a delay. The bounce message will typically include an explanation for the rejection, and the sender can then take appropriate action, such as resending the email. That is, this retry mechanism allows the SMTP server to deliver the email again. The responsibility for handling delayed emails ultimately lies with the SMTP server and its configuration. The representation of this mechanism aligns with the principles of Bernoulli feedback. Figure 16 shows a simple model of the components of the SMTP system.

8. Concluding remark

In this paper, we have studied on a single server with two types of general heterogeneous service under repeated service policy and Bernoulli feedback, which can be represented as $M/\binom{G_1}{G_2}/1$ queuing system. To make the queuing system into a Markovian process, we have introduced supplementary variables. The motivation for this model comes from wide range of applications arising in many real time systems encountered in various fields, such as in the manufacturing system, production system, telecommunication system etc. The queue size distribution at departure epoch, the response time distribution, inter-departure time distribution, and also the LST of the busy period distribution are derived. Further, this model is also observed by using the Embedded Markov chain technique and also obtain the PGF of the queue size distribution at departure epoch. The cost function of this model is also derived which is the most important in practical situation. Finally, the numerical illustration provides valuable insights into the system. In future, this analytic model can be generalized for T and D policy and also it can be observed for server breakdown and delayed repair as a future observation.

Acknowledgements

The authors would like to thank the editor and anonymous referees for their valuable remarks and suggestions which led to the paper in current form. The first author is grateful to Institute of Advanced Study in Science & Technology, a research institute, under DST, Govt. of India for providing the financial support.

References

[1] L. Abolnikov and A. Dukhovny, *Markov chains with transition delta-matrix: ergodicity conditions, invariant probability measures and applications*, Int. J. Stoch. Anal. **4**, 333–355, 1991.

[2] R.F. Anabosi and K.C. Madan, *A single server queue with two types of service, Bernoulli schedule server vacations and a single vacation policy*, Pakistan J. Statist. **19** (3), 331–342, 2003.

[3] A. Begum and G. Choudhury, *Analysis of an $M/\binom{G_1}{G_2}/1$ queue with Bernoulli vacation and server breakdown*, Int. J. Appl. Comput. Math. **9** (9), 1–32, 2023.

[4] P.J. Burke, *Delays in single-server queues with batch input*, Oper. Res. **23**(4), 830–833, 1975.

[5] B.D. Choi, B. Kim and S.H. Choi, *On the M/G/1 Bernoulli feedback queue with multi-class customers*, Comput. Oper. Res. **27** (3), 269–286, 2000.

[6] G. Choudhury and C.R. Kalita, *An M/G/1 queue with two types of general heterogeneous service and optional repeated service subject to servers breakdown and delayed repair*, Qual. Technol. Quant. Manag. **15** (5), 622–654, 2018.

[7] G. Choudhury and M. Paul, *A two phase queueing system with Bernoulli feedback*, Int. J. Inf. Manag. Sci. **16** (1), 35–52, 2005.

[8] D.R. Cox, *The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables*, in: Math. Proc. Cambridge Philos. Soc., Cambridge University Press, **51**, 433–441, 1955.

[9] M.M.N. GnanaSekar and I. Kandaiyan, *Nonlinear metaheuristic cost optimization and ANFIS computing of feedback retrial queue with two dependent phases of service under Bernoulli working vacation*, Int. J. Modern Phys. B, Doi: 10.1142/S0217979224400046, 2023.

[10] M.M.N. GnanaSekar and I. Kandaiyan, *Analysis of an $M/G/1$ retrial queue with delayed repair and feedback under working vacation policy with impatient customers*, Symmetry **14** (10), 1–18, 2024.

[11] M. Jain and S. Kaur, *(p, N)-Policy for unreliable server bulkqueue with Bernoulli feedback*, Int. J. Appl. Comput. Math. **6**, 1–28, 2020.

[12] C.R. Kalita and G. Choudhury, *Analysis of an unreliable $M/\binom{G_1}{G_2}/1$ repeated service queue with delayed repair under randomized vacation policy*, Comm. Statist. Theory Methods **48** (21), 5336–5369, 2019.

[13] I.E. Khan and R. Paramasivam, *Reduction in waiting time in an $M/M/1/N$ encouraged arrival queue with feedback, balking and maintaining of reneged customers*, Symmetry **14** (8), 1–18, 2022.

[14] B. Krishna Kumar, R. Rukmani, V. Thangaraj and U.R. Krieger, *A single server retrial queue with Bernoulli feedback and collisions*, J. Stat. Theory Pract. **4** (2), 243–260, 2010.

[15] S. Lan and Y. Tang, *Performance and reliability analysis of a repairable discrete-time Geo/G/1 queue with Bernoulli feedback and randomized policy*, Appl. Stoch. Models Bus. Ind. **33** (5), 522–543, 2017.

[16] K.C. Madan, A.D. Al-Nasser and A.Q. Al-Masri, *On $M^{[x]}/G_1G_2/1$ queue with optional re-service*, Appl. Math. Comput. **152** (1), 71–88, 2004.

[17] K.C. Madan, Z.R. Al-Rawi and A.D. Al-Nasser, *On $M^x/\binom{G_1}{G_2}/1/G(BS)/Vs$ vacation queue with two types of general heterogeneous service*, J. Appl. Math. Decis. Sci. **3**, 123135, 2005.

[18] S. Mahanta and G. Choudhury, *On $M/\binom{G_1}{G_2}/1/V(MV)$ queue with two types of general heterogeneous service with Bernoulli feedback*, Cogent Math. Stat. **5** (1), 1–9, 2018.

[19] S. Mahanta, N. Kumar and G. Choudhury, *An analytical approach of Markov modulated Poisson input with feedback queue and repeated service under N-policy with setup time*, Qual. Technol. Quant. Manag. **21** (2), 257–285, 2024.

[20] J. Medhi and J.G.C. Templeton, *A Poisson input queue under N-policy and with a general start up time*, Comput. Oper. Res. **19** (1), 35–41, 1992.

[21] A. Melikov, S. Aliyeva, S.S. Nair and B. Krishna Kumar, *Retrial queuing-inventory systems with delayed feedback and instantaneous damaging of items*, Axioms **11** (5), 1–17, 2022.

[22] S.P. Niranjan, S. Devi Latha, M. Mahdal and K. Karthik, *Multiple control policy in unreliable two-phase bulk queueing system with active Bernoulli feedback and vacation*, Mathematics **12** (1), 1–20, 2023.

[23] K. Rege, *On the M/G/1 queue with Bernoulli feedback*, Oper. Res. Lett. **14** (3), 163–170, 1993.

[24] L. Takács, *Introduction to the Theory of Queues*, Oxford University Press, New York, 1962.

[25] L. Takacs, *A single-server queue with feedback*, Bell Syst. Tech. J. **42** (2), 505–519, 1963.

[26] H. Takagi, *Queueing Analysis: A Foundation of Performance Evaluation*, Volume I: Vacation and Priority Systems, Elsevier Science Pub. Co. 1991.

[27] H. Takagi, *A note on the response time in M/G/1 queues wuth service in random order and Bernoulli feedback*, J. Oper. Res. Soc. Japan **39** (4), 486–500, 1996.

[28] S. Upadhyaya, *Performance prediction of a discrete-time batch arrival retrial queue with Bernoulli feedback*, Appl. Math. Comput. **283**, 108–119, 2016.

[29] R.W. Wolff, *Poisson arrivals see time averages*, Oper. Res. **30** (2), 223–231, 1982.